

# The Linked Open Data Wiki

## Empowering Organizational Knowledge Bases with Linked Open Data

SMWCon Fall 2017, Rotterdam, NL

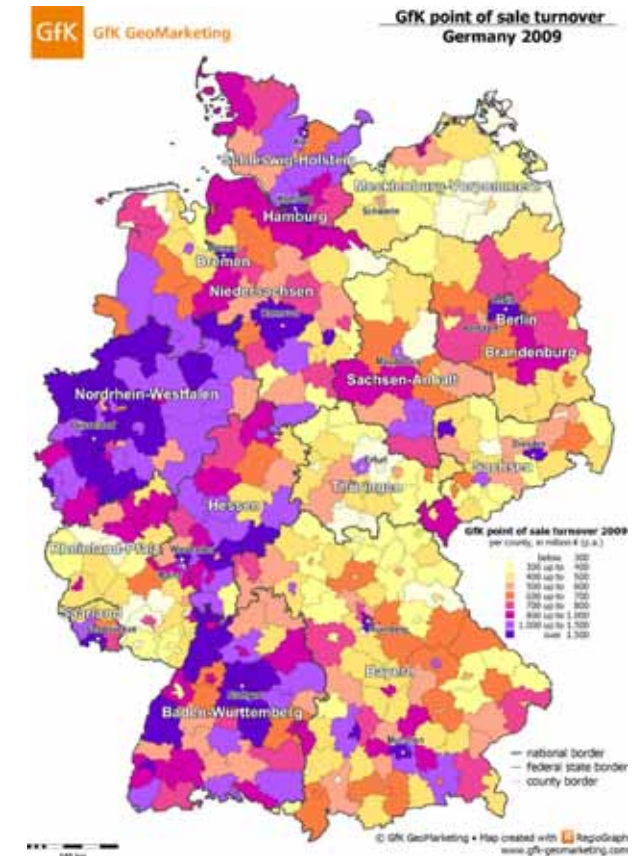
Matthias Frank, Research Scientist

05. Oktober 2017



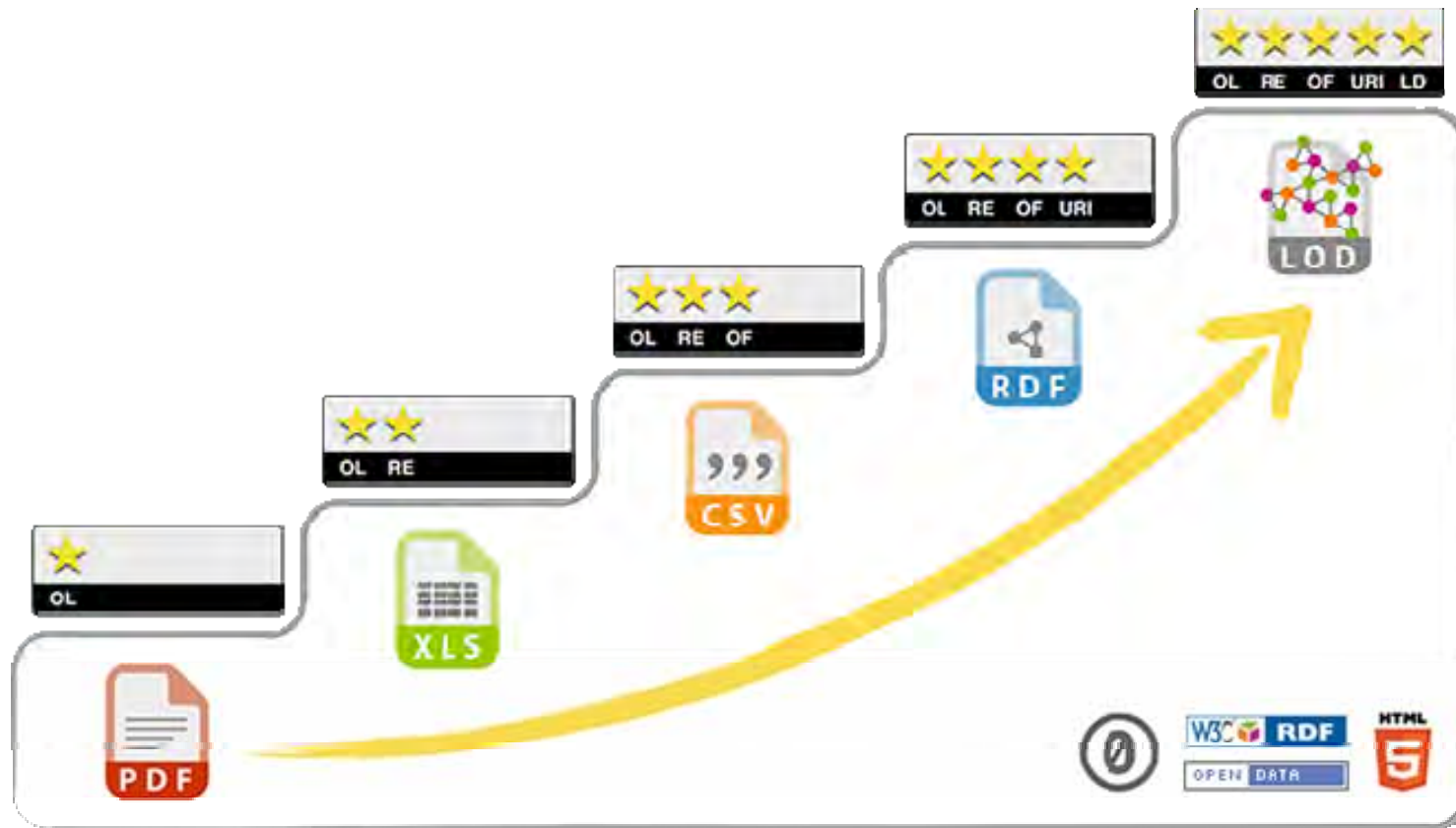
# Motivation

- Easy reuse of public available information for organizational knowledge bases
- *Example:* Knowledge base for geo a marketing campaign in SMB
  - *Public information:* population, coordinate location, area, elevation, sister cities, time zone, postal codes, coat of arms, local dialling code, licence plate code, ...
  - *Organizational knowledge:* customers, stores, employees, market survey, competition analysis, ...



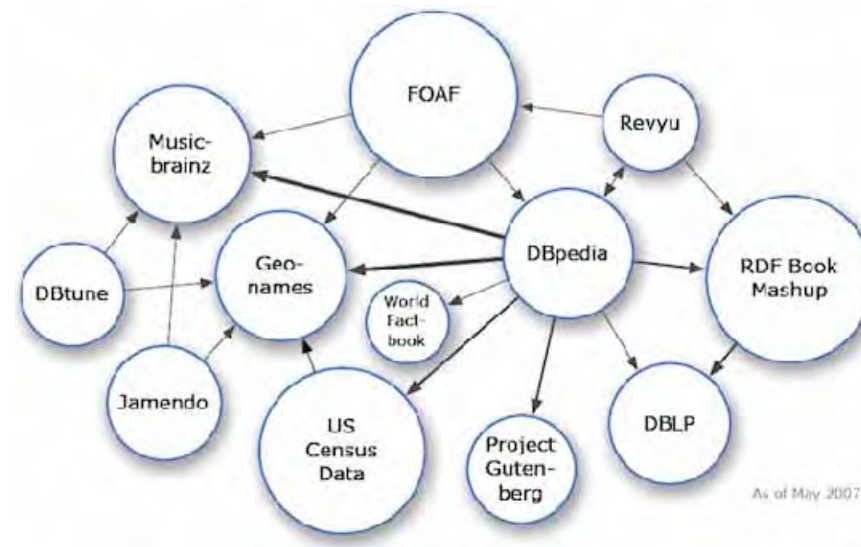
© GfK GeoMarketing GmbH

# Levels of Open Data



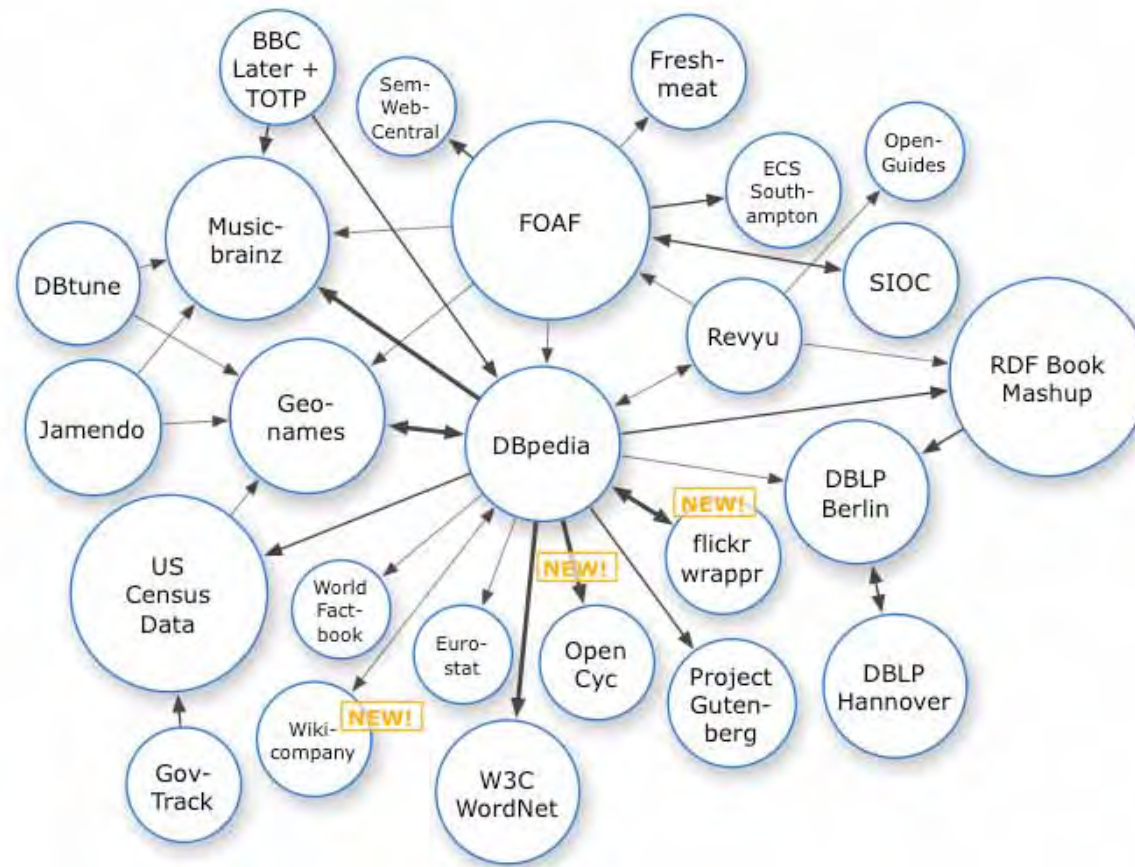
Berners-Lee  
(2006)

2007-05-01



## 12 Datasets

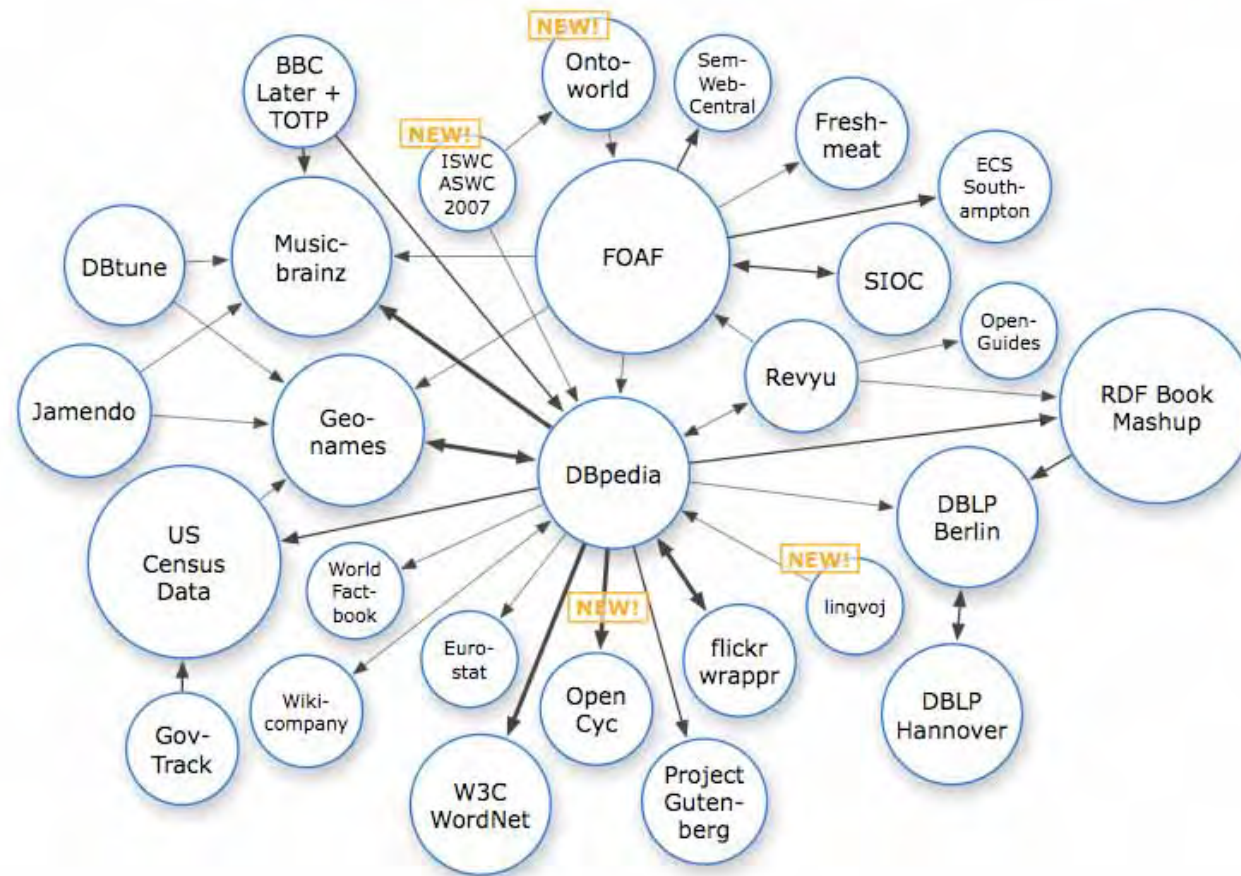
2007-10-08



25 Datasets

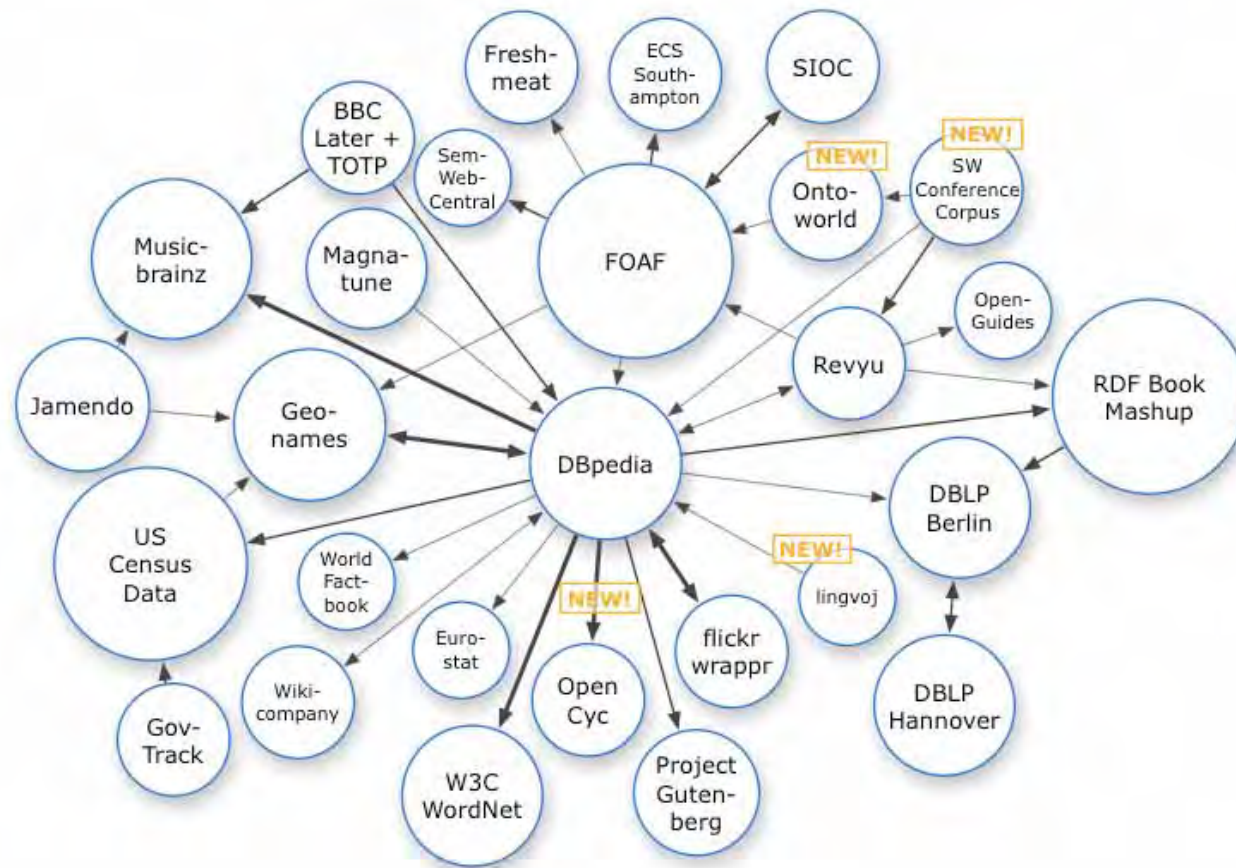


2007-11-07



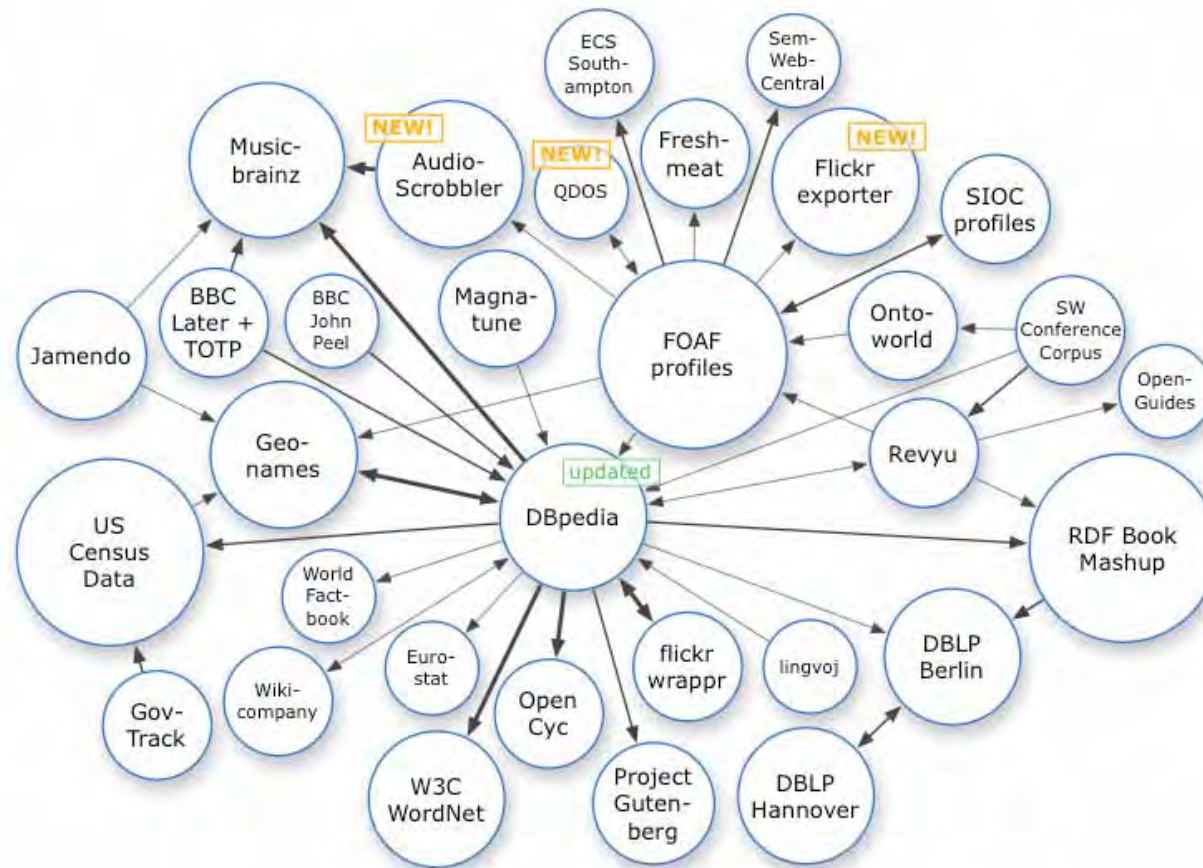
28 Datasets

2007-11-10



28 Datasets

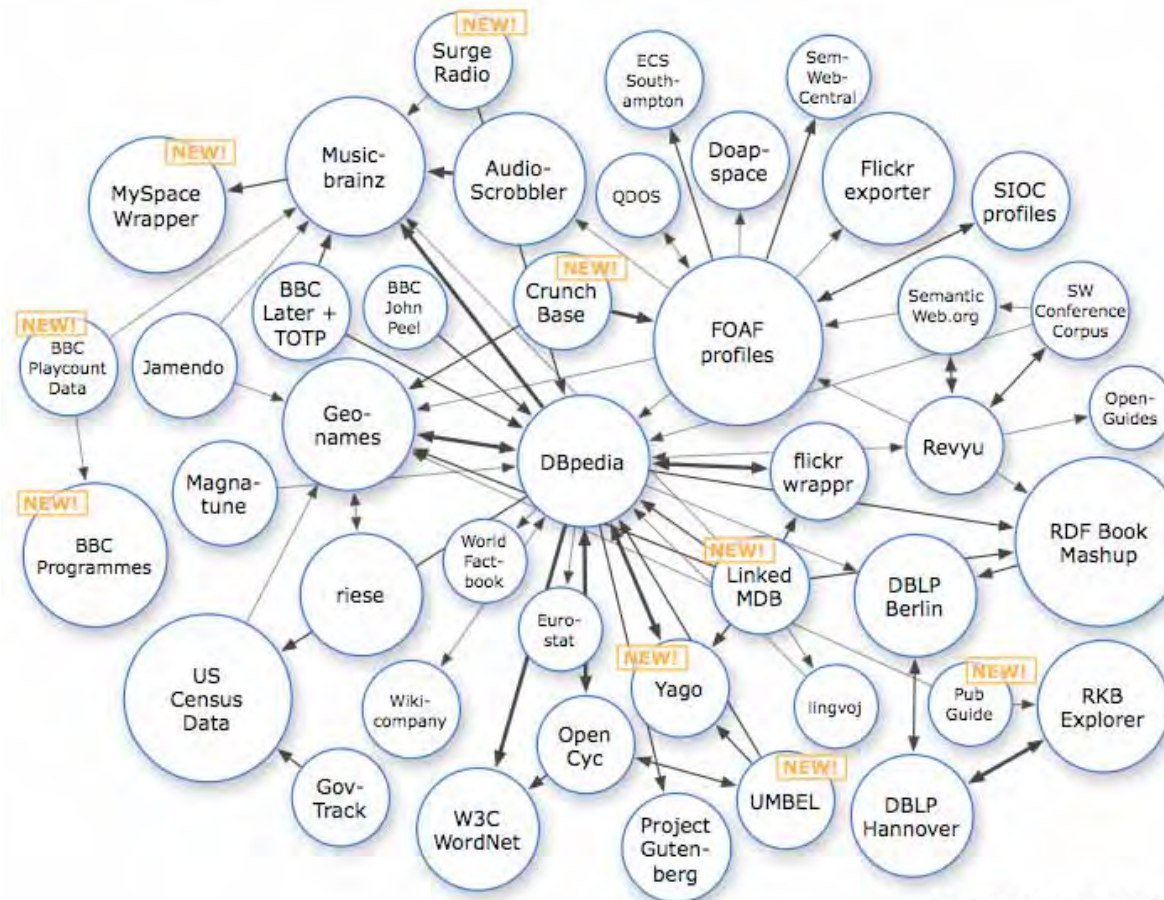
2008-02-28



32 Datasets



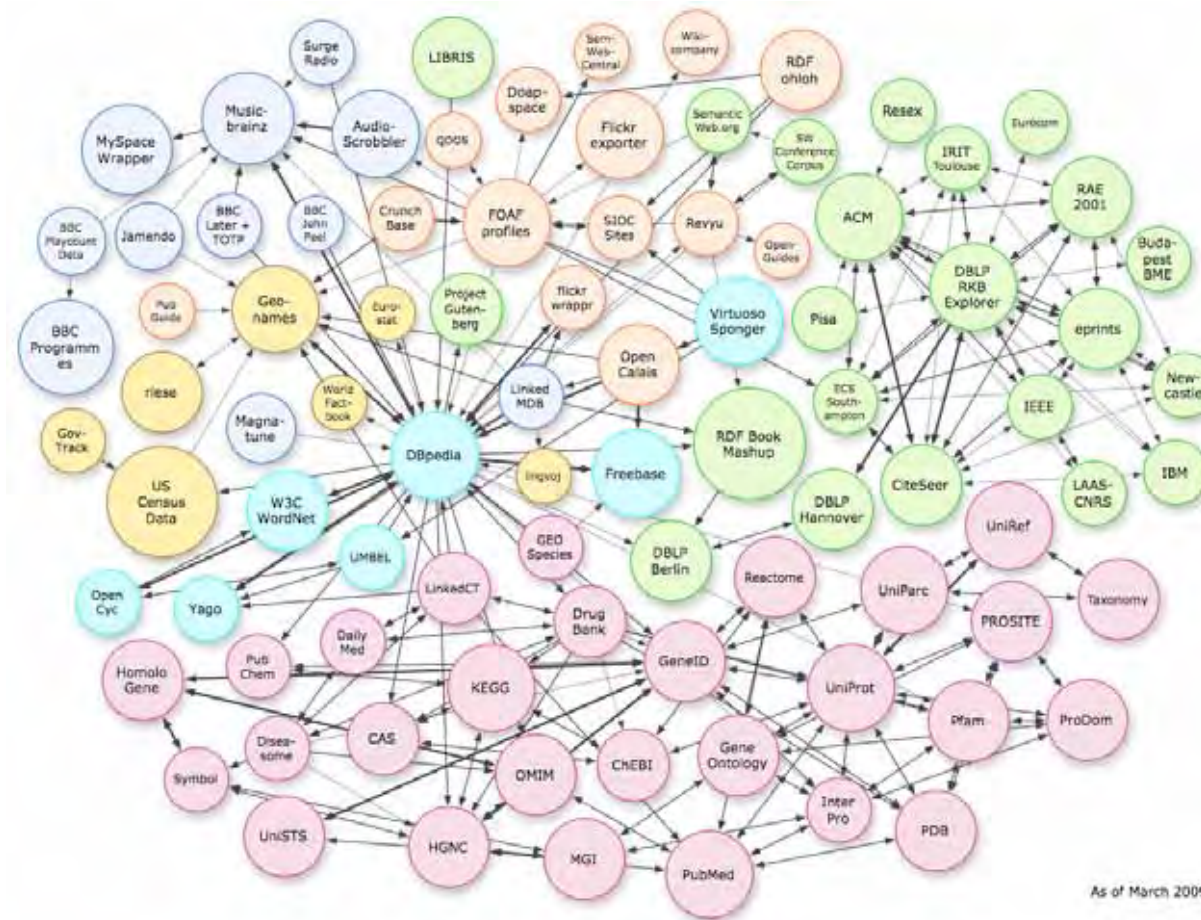
2008-09-18



As of September 2008

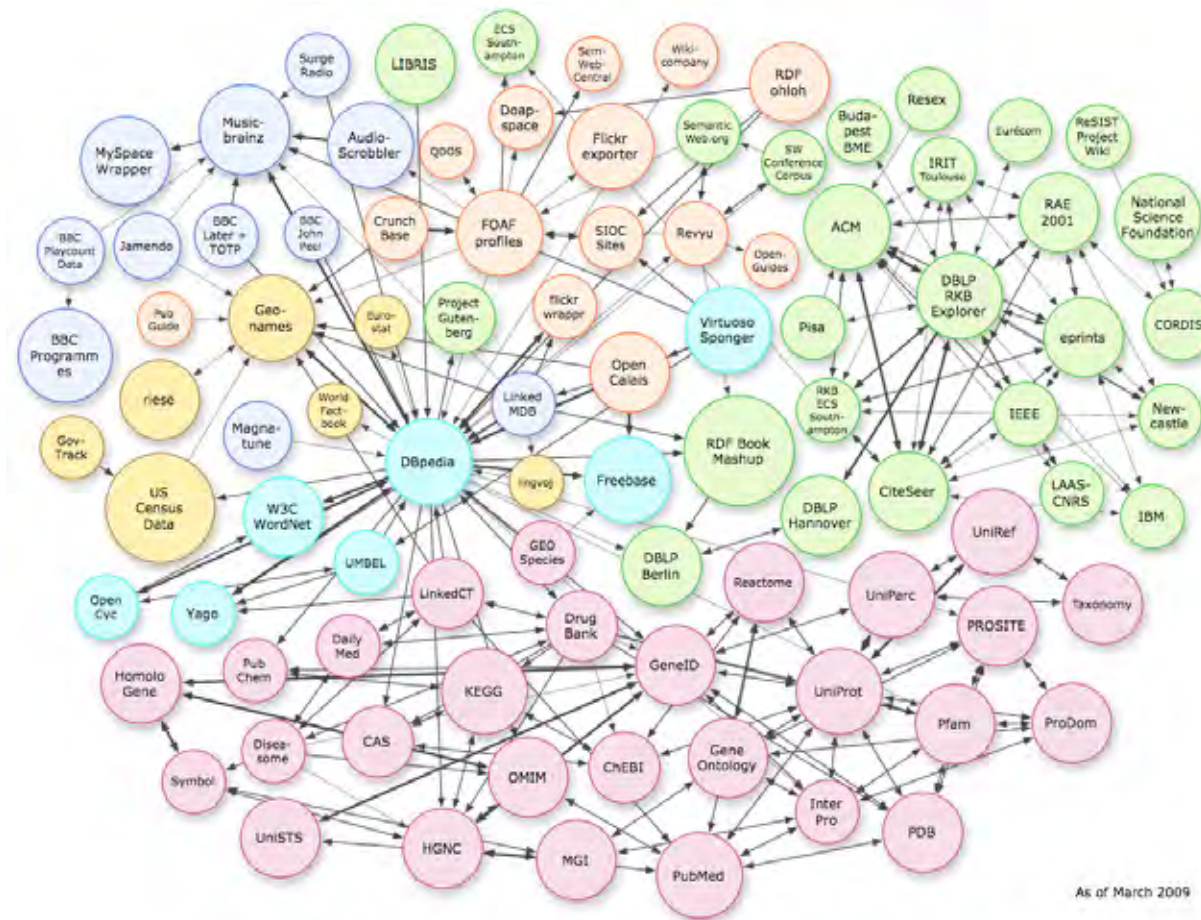
45 Datasets

2009-03-05



89 Datasets

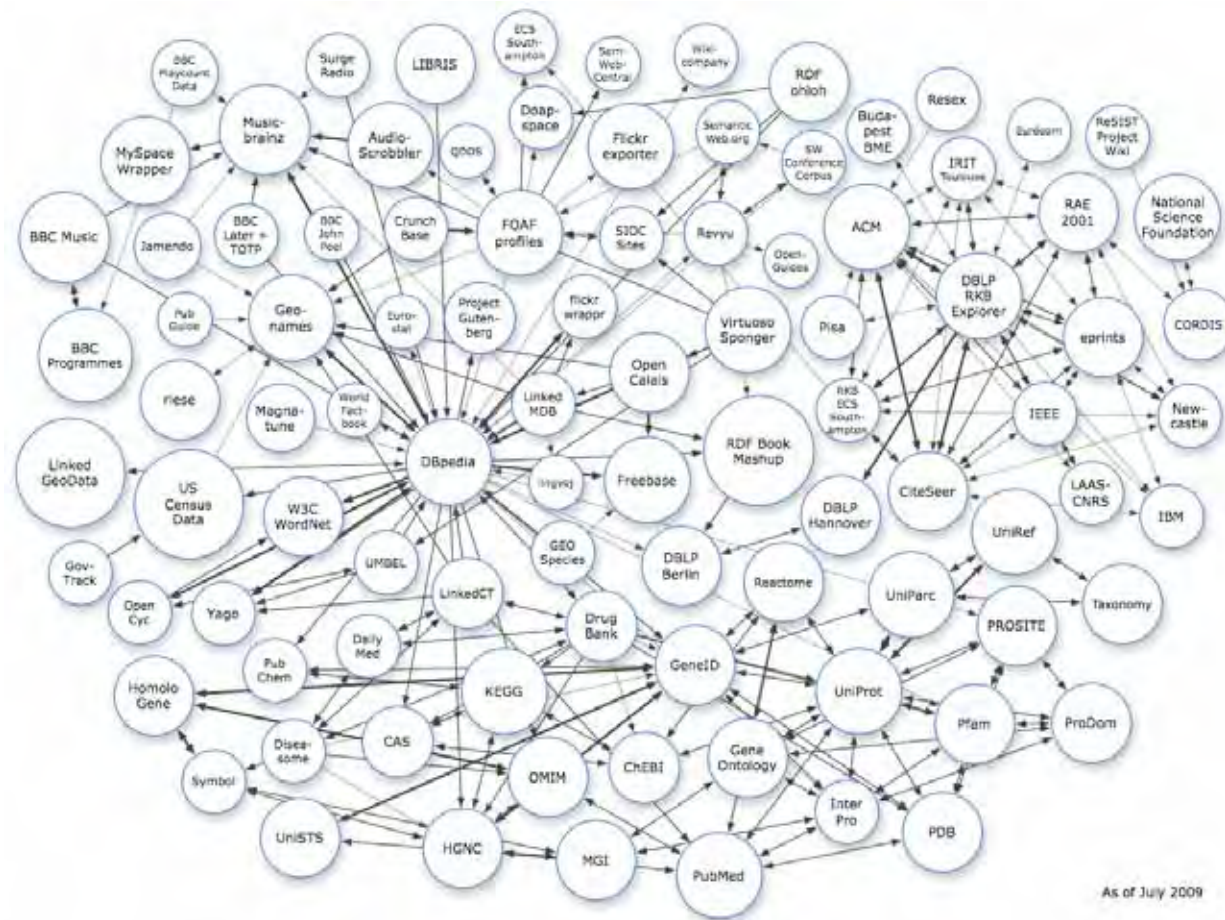
2009-03-27



## 93 Datasets

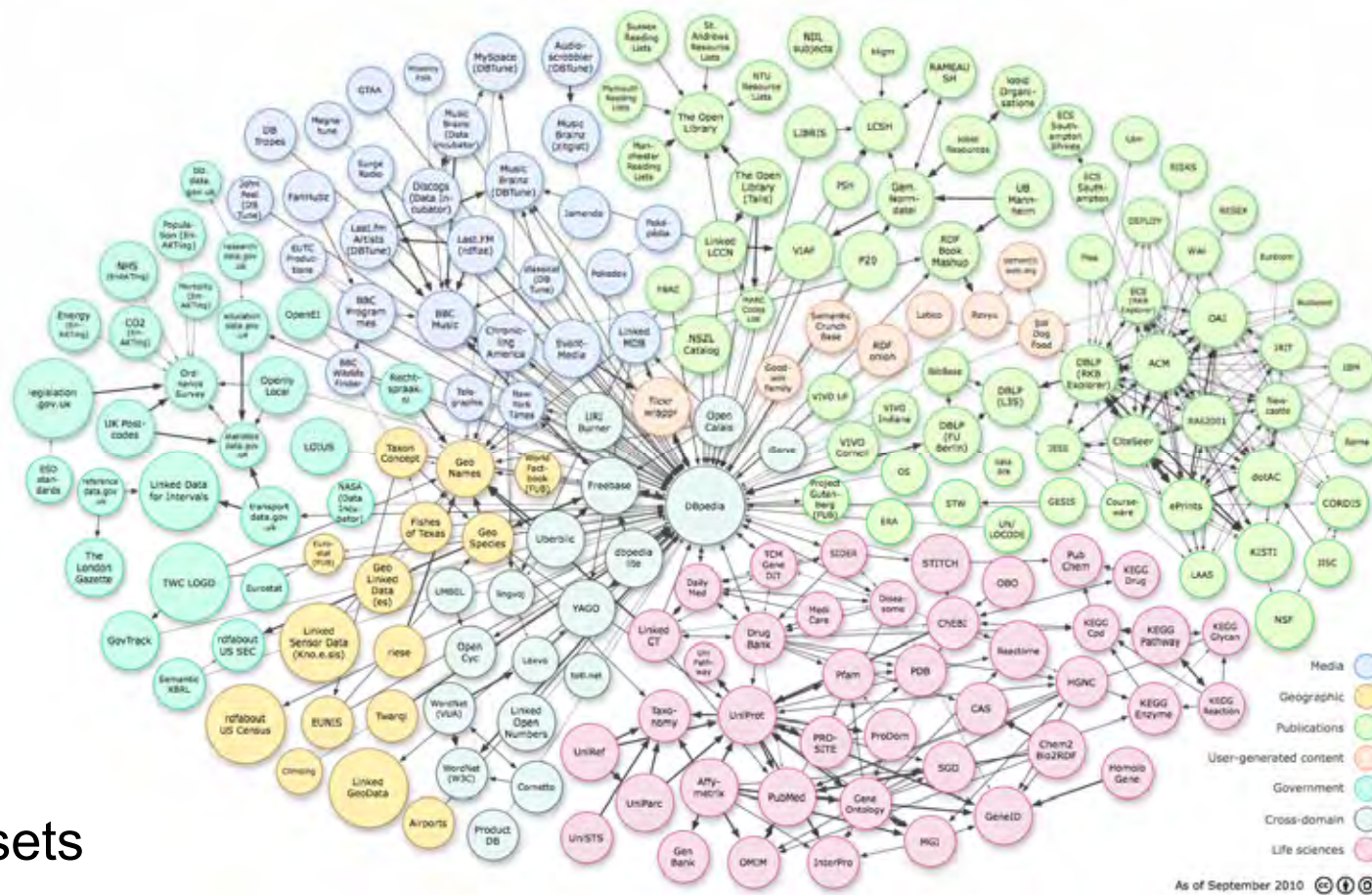


2009-07-14



# 95 Datasets

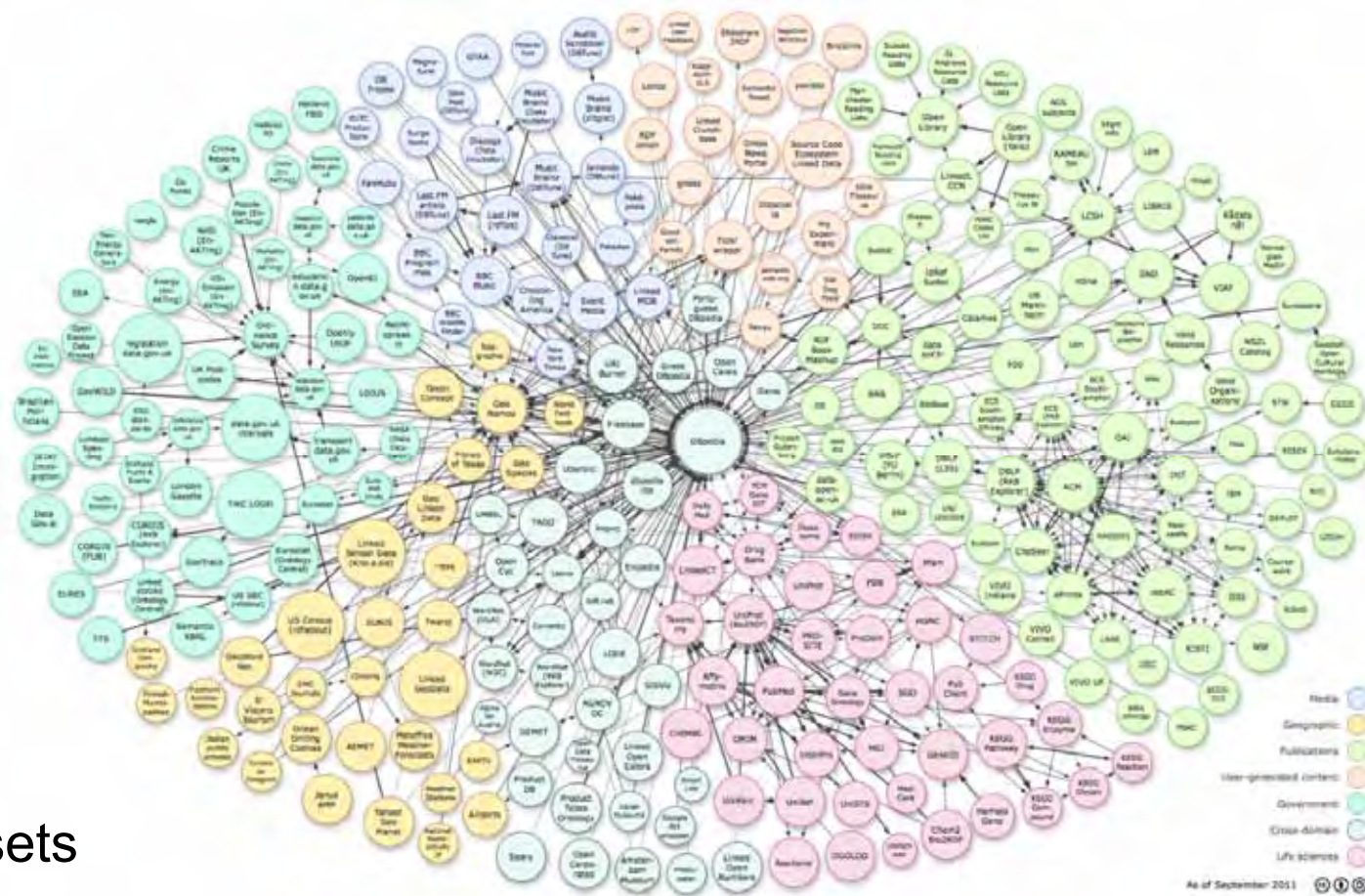
2010-09-22



203 Datasets

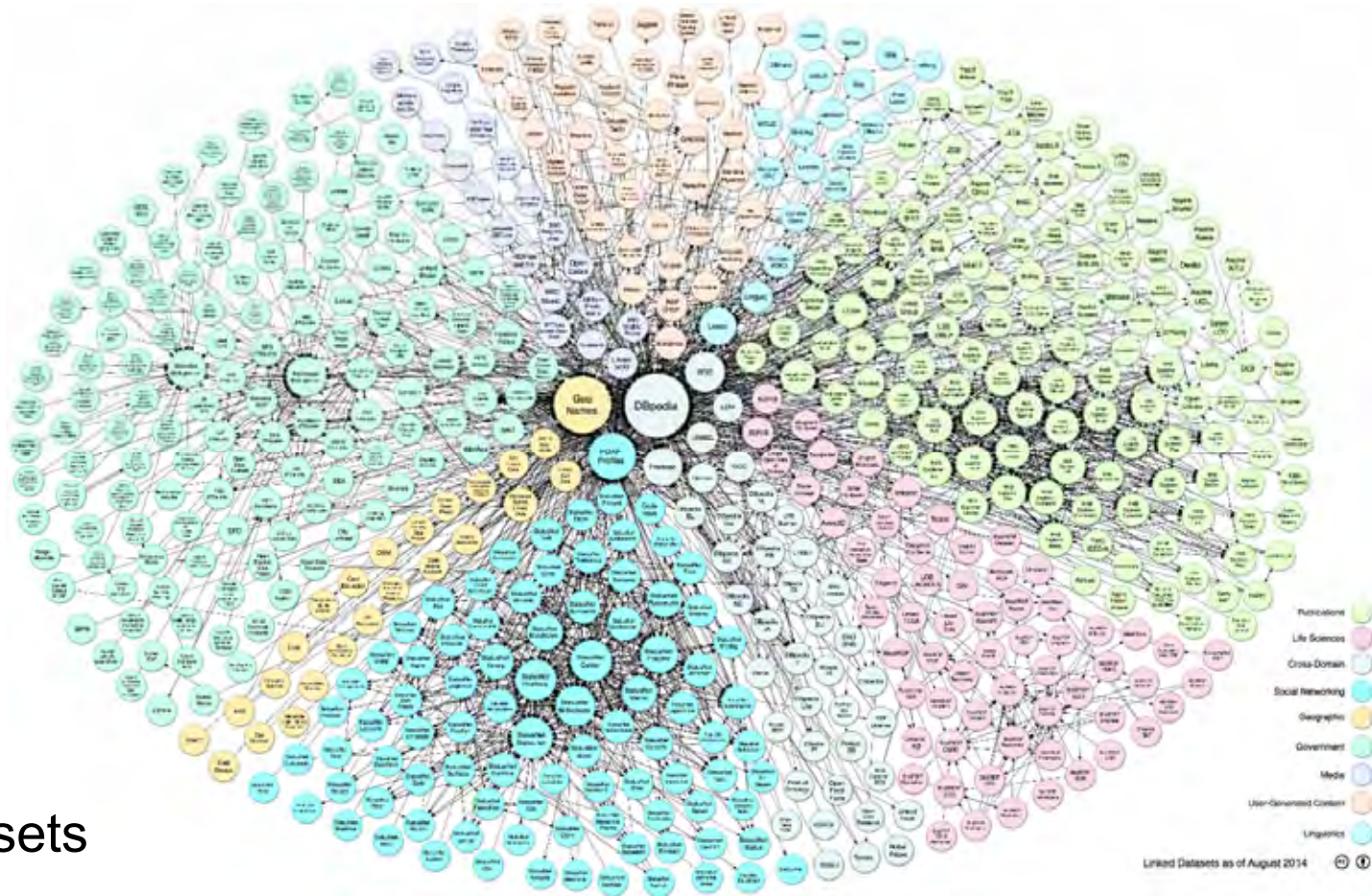


2011-09-19



295 Datasets

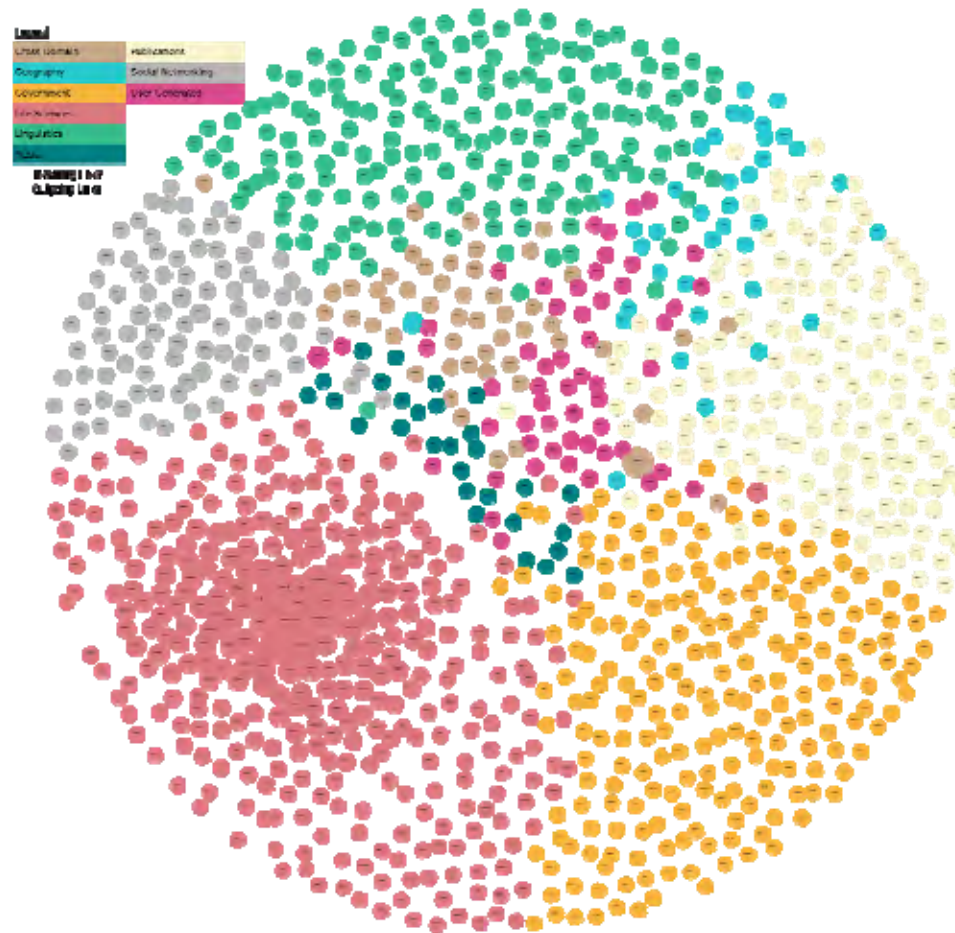
2014-08-30



570 Datasets

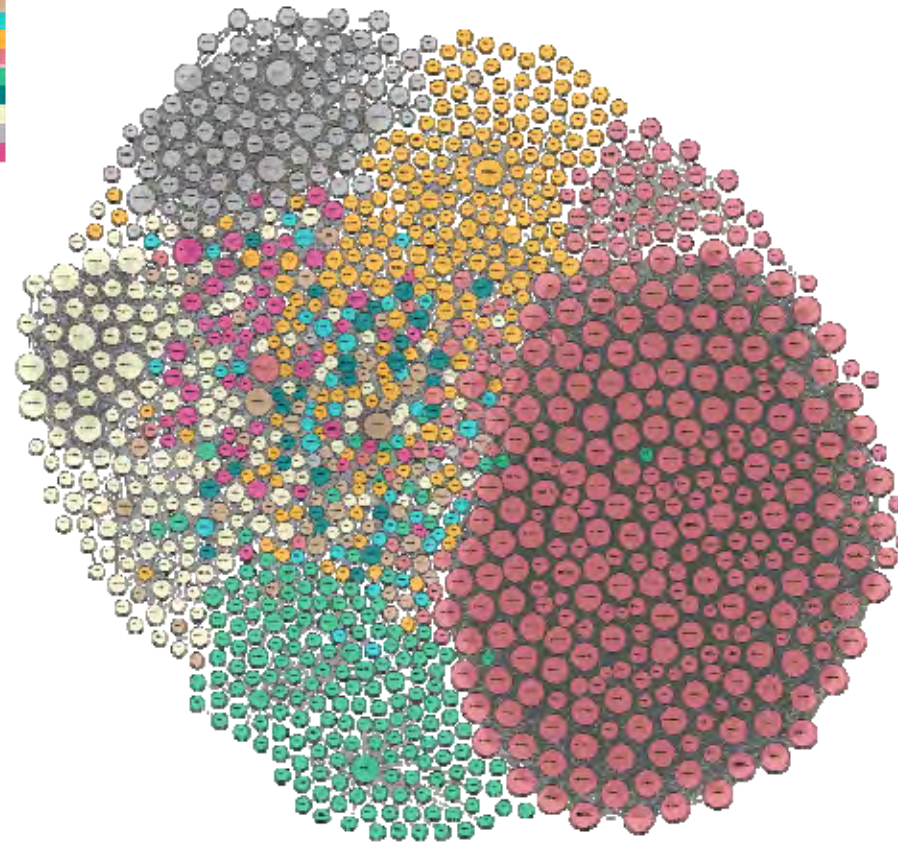


2017-01-26






1,146 Datasets

2017-02-20



1,139 Datasets

# Semantic Wiki Software (Open Source)

Approach	Underlying Engine	Data Storage / Export	Usage	Query construction	Integration of LOV/LOD
<b>MW / Cargo</b> 	MediaWiki	Relational, CSV export	Wikipedia, SMBs	#cargo_query (SQL-like)	-
<b>Semantic MW</b> 	MediaWiki	Relational, RDF mirror / export	Organizational knowledge management, e.g. Organizational knowledge graphs	#ask: (SPARQL)	manual import of single terms
<b>OntoWiki</b> 	-	Relational or RDF	Organizational knowledge graphs	SPARQL	publish ontology with LOV



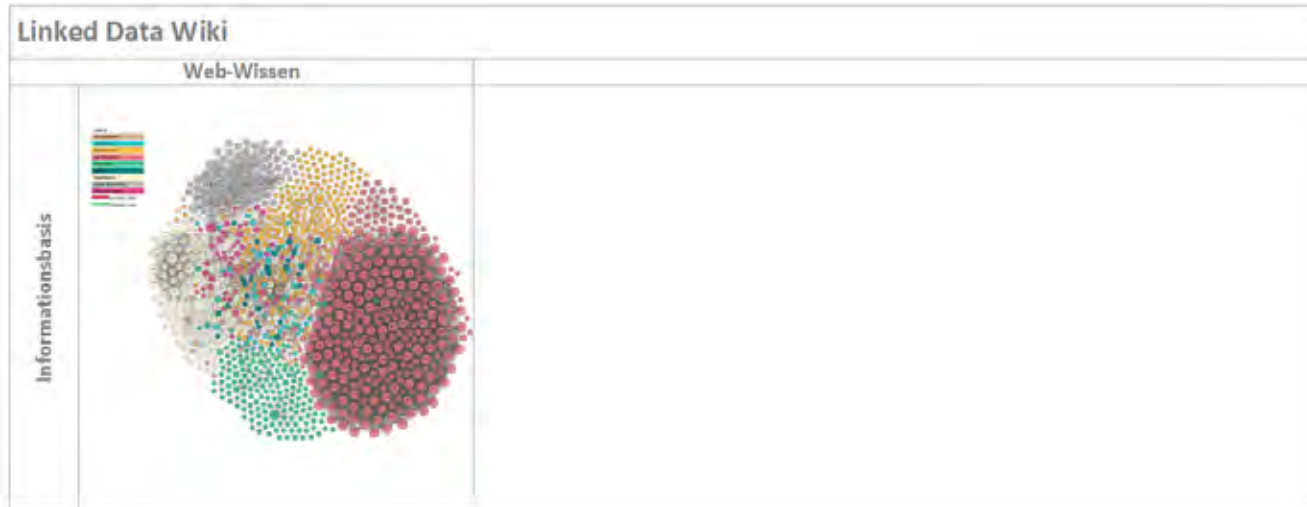
# Research Questions

- **RQ1:** How can we assist users of organizational wikis in establishing meaningful (i.e., semantic) relations to Linked Open Data entities?
- **RQ2:** How can we keep track of the provenance of statements in an organizational wiki, especially if these statements are inferred or gathered from Linked Open Data?
- **RQ3:** How can we evaluate and interpret potential uncertain, incomplete, inconsistent or redundant Linked Open Data correctly in order to increase the informative value of an organizational knowledge base?

# Approach

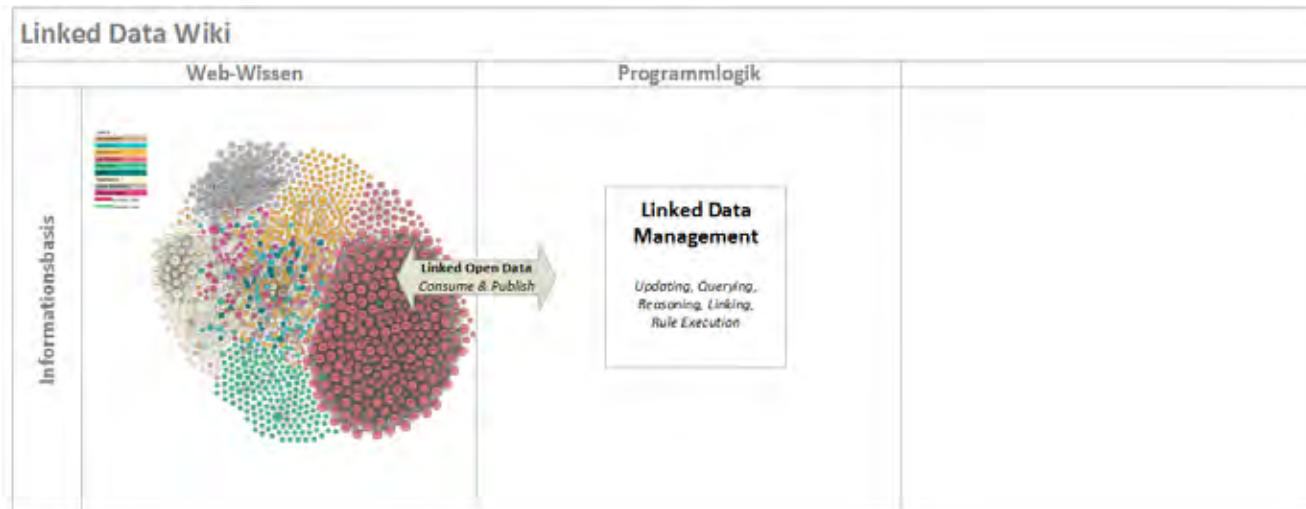
Linked Data Wiki

# Linked Data Wiki Architecture



**Web knowledge  
uses as information  
base**

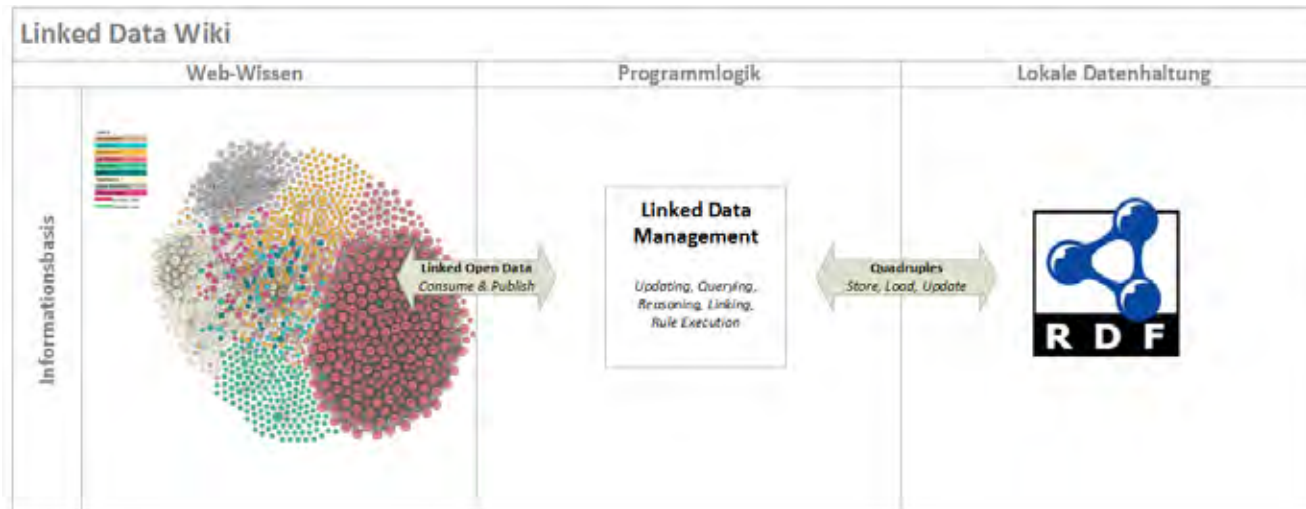
# Linked Data Wiki Architecture



## Linked-Data-Management

- Load from LOD
- Publish as LOD

# Linked Data Wiki Architecture

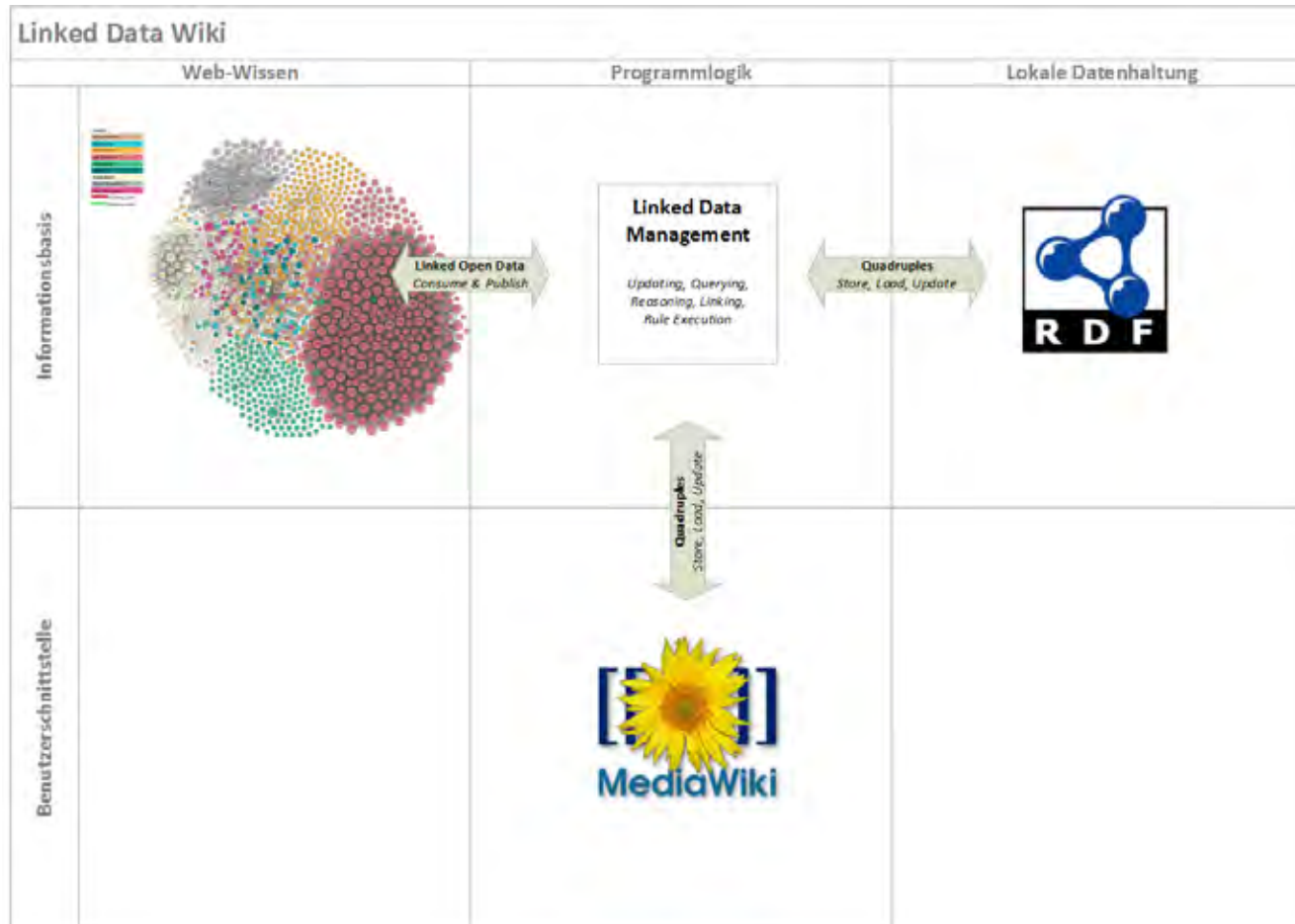


## Local Information

- Cache web knowledge
- Store additional information
- Load information
- Update information



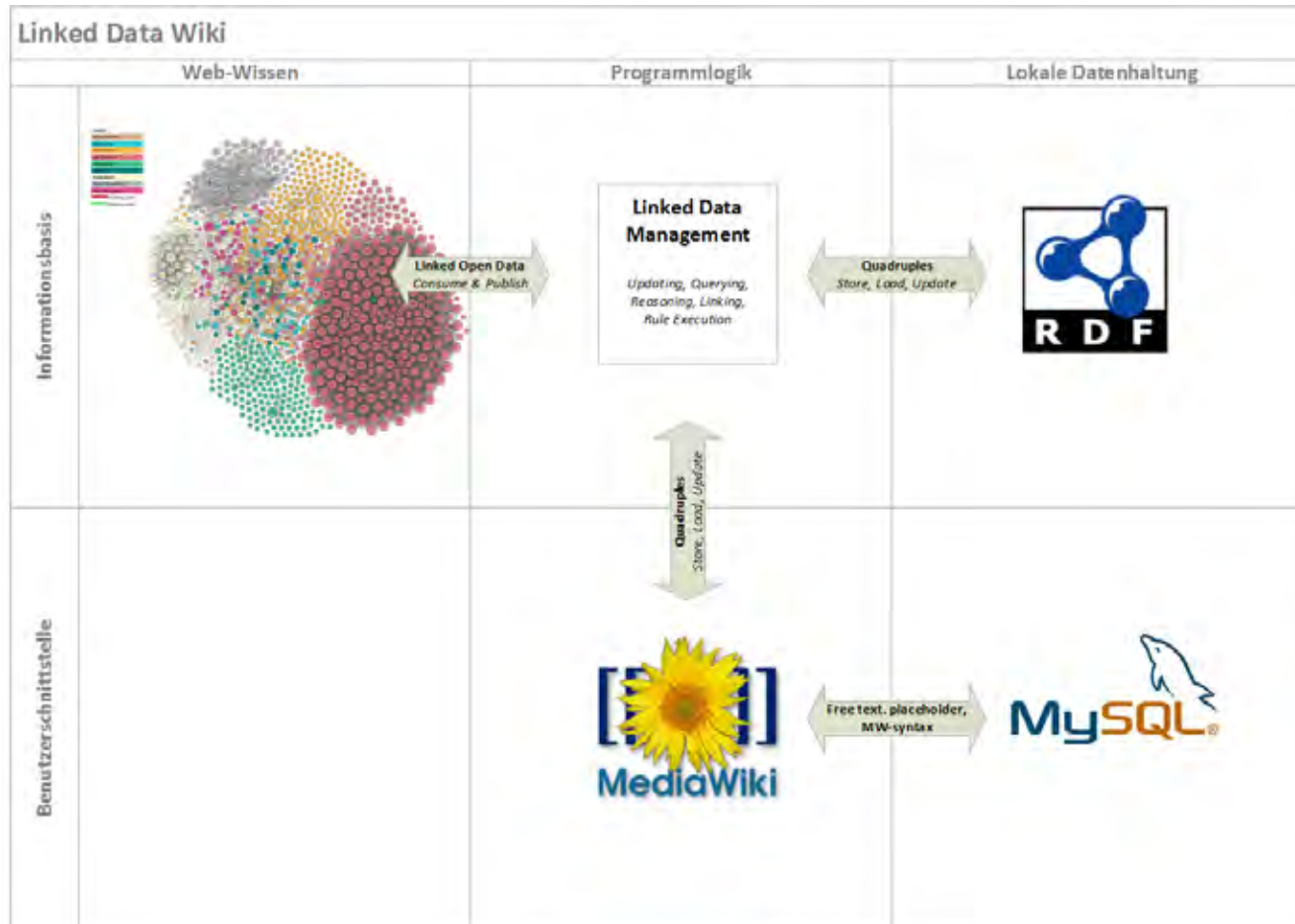
# Linked Data Wiki Architecture



## MW as user interface

- Load statements
- Add statements
- Update statements

# Linked Data Wiki Architecture



## Presentation data

- Free text
- Place holder
- MW-markup

# Implementation

# New Category

## Formular:Kategorie

Dies ist das Formular „Kategorie“. Um eine Seite mit diese automatisch zum Bearbeitungsformular der Seite weiterge

Stadt



```
SELECT * WHERE {
  ?category rdf:type rdfs:Class;
  rdfs:label "Stadt".
} limit 100
```



Spezialseite

### Kategorie erstellen: Stadt

Classes in LOD with label "Stadt": ☒ schema-org:City ☒ dbpedia-owl:City ☒ wikidata:Q515

Freitext:

- schema.org:City
- dbpedia-owl:City
- wikidata:Q515

Zusammenfassung:

☐ Nur Kleinigkeiten wurden verändert ☒ Diese Seite beobachten

[Datenschutz](#)
[Über SandBox](#)
[Haftungsausschluss](#)

# New Instance

Wikidata:Q1040

New page:  
“Karlsruhe”

Category: “Stadt”

```
SELECT * WHERE {  
  ?instance rdf:type  
    wikidata:Q515;  
    rdfs:label "Karlsruhe".  
} limit 100
```



The screenshot shows the Wikidata page for Karlsruhe (Q1040). It includes a table of language labels, a list of statements, and a list of references.

Language	Label	Description	Wikidata ID
English	Karlsruhe	German city in the state of Baden-Württemberg	
German	Karlsruhe	Stadtkreis in Baden-Württemberg, Deutschland	
French	Karlsruhe	ville de Bade-Wurtemberg	
Italian	Karlsruhe		

Statements

- instance of: [city](#) (wikidata:Q515)
- has city: [Karlsruhe](#) (wikidata:Q1040)
- located in the administrative territory of: [Germany](#) (wikidata:Q181)
- capital of: [Baden-Württemberg](#) (wikidata:Q1040)
- is a: [city](#) (wikidata:Q515)

References


- [Karlsruhe](#) (wikidata:Q1040)
- [Karlsruhe](#) (wikidata:Q1040)
- [Karlsruhe](#) (wikidata:Q1040)
- [Karlsruhe](#) (wikidata:Q1040)



# Schema knowledge

- Attributes of a city relevant for geo marketing:
  - population
  - coordinate location
  - area
  - elevation
  - sister cities
  - time zone
  - postal codes
  - coat of arms
  - local dialling code
  - licence plate code

# Identifying incomplete data

- Attributes of “Karlsruhe”:
  - population 296,033
  - coordinate location *49°0'50"N, 8°24'15"E*
  - area *173.46 square kilometre*
  - elevation *115 metre*
  - sister cities Nancy, Nottingham, Krasnodar, Timișoara, Halle
  - time zone *UTC+01:00*
  - postal codes 76229, 76131, 76137, 76133, 76135, 76139, 76149, 76199, 76185, 76187, 76189, 76227, 76228
  - coat of arms 
  - local dialling code
  - licence plate code KA

# Discussion

# Research Question 1

- How can we assist users of organizational wikis in establishing meaningful (i.e., semantic) relations to Linked Open Data entities?

## **Our contribution:**

- ✓ Query LOD automatically based on category and label

## Research Question 2

How can we keep track of the provenance of statements in an organizational wiki, especially if these statements are inferred or gathered from Linked Open Data?

### **Our contribution:**

- ✓ Store provenance information for each statement in the local repository using the Open Annotation Data Model

# Research Question 3

How can we evaluate and interpret potential uncertain, incomplete, inconsistent or redundant Linked Open Data correctly in order to increase the informative value of an organizational knowledge base?

## Ongoing Work:

- *Uncertain:*
  - Evaluate rules for data provenance
  - Compare values from different (independent) sources
- *Incomplete:*
  - Use schema knowledge from LOD to determine missing values
- *Inconsistent and redundant:*
  - Evaluate rules for data provenance
  - Use page ranks to get the most likely values





**THANK YOU**